DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

DETECTING FAILURES OF BACKWARD INDUCTION: MONITORING INFORMATION SEARCH IN SEQUENTIAL BARGAINING

Eric J. Johnson *
The Wharton School University of Pennsylvania

Colin Camerer
California Institute of Technology

Sankar Sen
School of Management
Temple University

Talia Rymon
Arison School of Business

SOCIAL SCIENCE WORKING PAPER 1040

June 1998

Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining*

Eric J. Johnson * Colin Camerer Sankar Sen Talia Rymon

Abstract

We ran three-round sequential bargaining experiments in which the perfect equilibrium offer was \$1.25 and an equal split was \$2.50. Subjects offered \$2.11 to other subjects, \$1.84 to "robot" players (who are known to play subgame perfectly), and \$1.22 to robots after instruction in backward induction. Measures of information search showed that subjects did not look at the amounts being divided in different rounds in the correct order, and for the length of time, necessary for backward induction, unless they were specifically instructed. The results suggest that most of the departure from perfect equilibrium is due to limited computation and some is due to fairness.

*The financial support of NSF 88-09299 and NSF 90-23531 to the first two authors is gratefully acknowledged. We also thank Rachel Croson, Brian Becker, Barry Blecherman, Gerard Cachon, David Goldstein, Teck-Hua Ho, Keith Weigelt, Ben Wilner, and many colleagues at Penn for their work and ideas. We also have received helpful comments from several referees and seminar participants at many universities including Harvard, Cornell, New York University, Penn State, Rochester, Toronto, Minnesota, MIT, and the International Conference on Game Theory (Florence) and BoWo IV (Bonn).
*E-mail: johnson00@wharton.upenn.edu.

1. Introduction

Game-theoretic models of sequential bargaining, in which agents alternate offers and counteroffers, have proved useful in economic theorizing in several ways (e.g., Ståhl, 1972; Rubinstein, 1982). The models serve as building blocks for theories of imperfectly competitive markets, such as "matching markets" with small numbers of buyers and sellers (see Osborne and Rubinstein, 1990). Sequential bargaining models with incomplete information have been useful in explaining facts about labor strikes (Kennan and Wilson, 1990). Alternating-offer models also give a procedural underpinning to axiomatic cooperative solution concepts such as the Nash bargaining solution; (Binmore, Rubinstein, and Wolinsky, 1986), unifying different styles of analysis.

Perhaps surprisingly, direct experimental tests have typically rejected the predictions of sequential bargaining models. Actual offers are closer to equal divisions of an initial "pie size" than to equilibrium offers, and offers do not change with parameter changes as theory predicts (see Roth, 1995; Camerer, in progress).

An example will illustrate the game we use in our experiments, and the typical result. Two players, 1 and 2, bargain over a pie that shrinks in value.[1] in each of three periods (reflecting discounting due to impatience). Suppose the pie is worth \$5 in the first period, \$2.50 in the second period, \$1.25 in the third period, and nothing after that. Starting with player 1, players alternate making offers, which the other player can accept or reject (moving to the next period).

Assuming players are purely self-interested, perfect equilibrium divisions can be derived by backward induction, starting with the third-period division and working backward. The perfect equilibrium is for player 1 to offer \$1.25 and keep the rest for himself, give or take two cents[2].

In most experiments, subjects actually offer something between the \$1.25 equilibrium and \$2.50 (which is an equal split of the first-round pie). Offers average around \$2.00. Lower offers, including offers near the equilibrium prediction of \$1.25, are frequently rejected.

1.1 Two theories: fairness and limited computation

In this paper, we compare two theories of why these observed offers occur, which we label "fairness" and "limited computation" (or bounded rationality).

Fairness theories account for departures from perfect equilibrium by positing a "social utility" for others' payoffs and aversion to inequality (Loewenstein, Thompson and Bazerman, 1989; Bolton and Oeckenfels, 1997; Fehr and Schmidt, 1997) or a preference for reciprocating fairness and unfairness (Rabin, 1993; Levine, 1997). The clearest evidence of fairness comes from dictators and ultimatum games (see Camerer and Thaler 1995), in which offers are much more generous than predicted by equilibrium. Fairness theories take the average \$2 offer in our three-period example as evidence of a failure of the assumption of pure self-interest. The inequality-aversion and reciprocity theories assume players are reasoning game-theoretically, but care about fairness, while other accounts leave open the possibility that reasoning is not game-theoretic (e.g., Güth and Tietz, 1990).

Limited computation theories explain observed departures from perfect equilibrium by suggesting that players do not reason game-theoretically, and hence do not initially understand how the structure of the game conveys bargaining power. Confused or naive, players offer equal splits because they don't know what else to do. According to this view, fair offers (close to \$2.50) are only temporary-- fairness is a heuristic for generating an offer, not an expression of preference-- and will be displaced by strategic behavior as suitable experience accumulates. Evidence consistent with this view comes from experiments in which special kinds of experience lead to equilibrium offers after subjects "live through" bargaining subgames (Binmore, Shaked and Sutton 1985; Harrison and McCabe, 1992[3]). With this experience, subjects' offers sometimes converge away from equal splits and toward the subgame perfect equilibrium, although most experiments show little learning of this sort (Ochs and Roth, 1989; Spiegel, et. al. 1994).

Combining elements of fairness and limited computation yields four possibilities, which are listed in the first column of Table I. The table shows the offers these theories predict subjects will make in different experimental treatments. We first focus on the left column of Table I, the predictions theories make when players bargain with other subjects.

If players care only about their own payoffs, and can compute perfect equilibrium, then the game-theoretic prediction (GT) of $1.25 results. Now suppose players can compute perfect equilibrium but are unsure of the rationality of others, or are constrained by fairness, either because they prefer to treat other subjects fairly, or believe that others may reject unfair offers. These distinct possibilities all make the same predictions across our experimental treatments, which we subsume under the label "game-theoretic with fairness" (GT-F). These theories predict offers will lie in an interval between the GT prediction and the equal-split point, [$1.25,$2.50].

Now suppose players do not reason game-theoretically and do not initially understand how the structure of the game creates bargaining power. Under this "limited computation" (LC) theory, predicted offers could lie anywhere; if we assume an upper bound at the equal split point $2.50, we predict an average offer p_{LC} somewhere in the range [0,2.50]. If players have limited computation <u>and</u> are concerned with fairness, they will offer an amount $p_{LCF}(s)$ to other subjects, which will be larger than that made under simple limited computation, hence $p_{LCF}(s) \geq p_{LC}$.

Insert Table I here

Previewing the results of our studies, the bottom line of Table I shows that the average offer our subjects made when bargaining with other subjects was $2.11. This offer rejects the GT prediction of $1.25 but does not enable us to tell whether players are reasoning game-theoretically but are constrained by concerns about rationality or fairness (GT-F), or are not reasoning game-theoretically (LC or LC-F). We use two additional treatments to distinguish these two classes of explanations.

In one study, subjects made offers to "robot" players who they know will play subgame perfectly and have no concern for fairness. The robot condition removes the influence of fairness-- because no other person receives the offers subjects make, and the robots don't care about being treated fairly-- and sharply penalizes deviations from perfect equilibrium behavior.[4] If players are game-theoretic (GT) or game-theoretic but are constrained by fairness (GT-F), they will offer only $1.25 to the robots. Under the limited computation theory LC, players will make the same offer p_{LC} to robots that they made to other human players (since they don't know how to compute perfect equilibrium offers). Under LC-F, players will make smaller offers to the robots, denoted $p_{LCF}(r)$, than the $p_{LCF}(s)$ they offered to other subjects, presuming they care less about treating robots fairly. The key prediction of the limited computation theories is that subjects do not instinctively play game-theoretically, but could perhaps they could learn to do so. Our third "instructions" treatment allows this learning in a novel way, by instructing the players in the logic of backward induction, then having them play robots afterward. If the observed discrepancy between the game theoretic prediction and actual offers is due to a lack of understanding of backward induction, then instruction should lead to more subgame perfect play. All theories predict that offers should be $1.25 in the "robots after instruction" condition. LC theories can be distinguished from GT theories because they predict that equilibrium offers will <u>only</u> be made in this condition.

The average offers in all three conditions, shown in the bottom line of Table I, support the view that subjects use limited computation <u>and</u> are constrained by fairness (LC-F). Average offers to robots are $1.84, and $1.22 after instruction. The fact that subjects offer less to robots than to other subjects implies that they are partly constrained by fairness toward other people. The fact that they make the perfect equilibrium offer only after instructions means backward induction seldom occurs naturally in these games. Note that the GT-F prediction, including variants in which players are game-theoretic and rational but are not sure others are, is rejected because game-theoretic subjects should offer the robots $1.25 and they do not.

2. Studying Cognition with Measures of Information Acquisition

The mean offers across the three conditions support the limited computation view, but the offers alone do not help us understand what computations subjects are doing. To explore the nature of cognition further, we recorded the information subjects looked at, and for how long, as they bargained. These information-gathering data illuminate the heuristics subjects are using and enable an indirect test of whether they are computing perfect equilibria.

Insert Figure 1 here

Figure 1 shows the computer display subjects saw, created using a software system called MOUSELAB which has been used in individual choice research.[5] The computer screen has 6 boxes. Hidden behind each box is the amount of the pie to be split in a round (left-hand column boxes) or the role of the subject in a round (right-hand column boxes), for each of the three rounds of the game. To see what is in a box, subjects used a mouse to move a cursor into the box. Once the cursor enters the box, the box automatically opens and reveals the contents of the cell; the box closes when the cursor is moved outside it. By recording which box is open every 16.67 milliseconds, we collect a fine-grained measure of the time when each box is entered and exited, allowing us to study the order in which boxes were opened, and how long each box was opened. In Figure 1 the cursor is in the upper left-hand box, so the player sees that the pie in the first round is $5.00.[6]

Players make an offer by moving the cursor to the dollar scale at the bottom of the box. As the arrow moves along the scale, the numerical value the arrow indicates is shown in a box to the left of the response scale ($1.29, in Figure 1).

We use information acquisition to test game theoretic principles as if they are predictions about cognition. This runs counter to an important convention in economics--theories should be judged only by their implications, because models can make good predictions even if their assumptions are false. However, game-theoretic sequential bargaining models <u>do not</u> make good predictions about initial offers. Finding out which underlying assumptions are false is the next logical step, and direct tests of information search are one way to do so.[7]

An analogy can be made to revealed preference. To study unobservable preferences, we often ask subjects to choose between objects (commodity bundles, gambles) which reveal their preferences. Similarly, asking subjects to `choose' information provides clues to unobservable thinking patterns. Just as educated guesses about a factory's production process can be made by observing the flow of inputs, length of processing time, and final output, the information search data can be used to make educated guesses about how subjects are thinking. For example, if a subject does not open the third box at all, showing the third-stage pie size, she cannot be computing a perfect equilibrium offer.

Of course, game theory makes no specific prediction about exactly how information is processed to reach a perfect equilibrium. However, we can construct an empirical benchmark for game-theoretic information gathering by observing information-gathering of subjects instructed in how to do backward induction. Since these subjects actually do make equilibrium offers, we take their patterns to be a picture of what information search looks like under game-theoretic thinking. The next section shows this picture. If the information-gathering of regular subjects is different than the game-theoretic picture, then regular subjects are not thinking game-theoretically.

3. Study 1: Calibration

Before turning to our two main studies, we describe a calibration study which illustrates our general methodology and shows the kind of information acquisition patterns exhibited by subjects trained to use backward induction.

As in all the following experiments, subjects were Wharton undergraduates recruited from business or economics classes, or from general sign-up sheets posted around the campus.[8] Ten students met in our lab at a designated time. An experimenter read instructions (see Appendix A) aloud as the players read their own copies, to make the instructions public knowledge. Afterwards, they worked through some examples (balanced to avoid biasing the results) and took a quiz to ensure they understood the instructions.

In this calibration session, players only made first-round offers; they did not actually bargain with others. In all the trials, subjects were told to generate offers that would maximize the amount of money they would get from bargaining with a self-interested, profit-maximizing opponent. They were paid (at the end of the session) $0.50 for each trial in which their answer fell within $.05 of the equilibrium offer. We varied the instructions halfway through the experiment: We first ran 8 trials with the standard instructions (Appendix A), then 8 more trials after additional instructions in how backward induction could be used in these games (see Appendix B). The instructions did not explicitly mention how to look for information. Two examples were worked on the blackboard and questions were answered.

3.1 Results

The calibration-study data address two questions: First, can backward induction instructions produce subgame perfect equilibrium play? The answer is Yes. The average offer before instruction, in the first eight trials, is $1.84. After receiving instruction, the average is $1.22 in the last eight trials, which is not significantly different than the equilibrium prediction.[9] The difference between the two sets of trials is highly significant, and there is no significant effect of experience within each group of trials.

Second, are the changes in offers accompanied by changes in information acquisition? We examine three measures of information search: Number of acquisitions per box (the number of times each box is opened in a period); total time examining payoff, or "looking time", per box (the amount of clock time each box is open in a period[10]); and the number of transitions from one specific box to another.

Before presenting these data, it is useful to ask: Suppose we take perfect equilibrium to be a computational algorithm. What does equilibrium analysis predict about the pattern of information search by subjects? Player 1s calculating equilibrium offers might search for information as follows: They first look at the third round payoff, and calculate that round's equilibrium offer. Then they look at the second round payoff, and figure out that round's equilibrium offer. This may take more time, since subjects may glance back and forth between the second- to the third-round boxes, to see the pie sizes and calculate their difference. Finally, they may look at the first round payoff to see how much remains.[11] Thus, we take equilibrium analysis to predict: Transitions from box to box will be predominantly backward transitions, from the nth-round box to the n-1st round box; and the longest looking time will be in the second box.

To depict information acquisition data we use a trio of icon graphs, a multivariate display representing the three acquisition measures (see Figure 2). Each graph corresponds to one of the three payoff boxes on the computer display. The width of each box is proportional to the number of acquisitions of that box and the height of the shaded area in each box is proportional to the amount of time spent looking at that box. Each of these measures is standardized, so that the widest or most shaded box is the one which is most acquired or open longest (in Figure 2, that is the second round payoff in the third column). Horizontal lines show box midpoints to facilitate comparison. The arrows represent the number of transitions between boxes, with the thickness of each arrow proportional to the transition frequency. Transitions which occurred less than once a trial, on average, are omitted to simplify the display. The left column of Figure 2 shows the hypothetical pattern corresponding to our description of information acquisition predicted theoretically by backward induction.

Insert Figure 2 here

The middle and right columns of Figure 2 represent actual results from the control trials (rounds 1-8) and instructional trials (rounds 9-16). The icon graphs show dramatic differences in how subjects process information before and after instruction. During the control trials, subjects spend much less time looking at the second round payoff and almost none looking at the third. After backward induction instruction, subjects look at the third and second rounds most often (and move back and forth between them frequently), and spend almost no time looking at the first round box, a pattern which is consistent with our hypothetical characterization of backward induction. Because instruction leads subjects to generate equilibrium offers, and search for information in the manner we conjectured, we conclude that (1) the instructions do teach the backward induction strategy for calculating the equilibrium and (2) monitoring the pattern of information acquisition can detect backward induction.

4. Study 2: Bargaining with other players

Study 1 indicates how the process of backward induction is manifested in measures of information search. It does not address whether or not backward induction occurs in actual bargaining between people. To examine this question we ran multiple sessions of an alternating-offer game.

Each group of ten subjects played eight three-round alternating-offer bargaining games, with a different anonymous opponent each time. This general design helps subjects to learn from "stationary replication" of the game, while avoiding the reputation effects which might arise if two subjects knew they were playing each other repeatedly. At the end of the session they were paid at a rate of $.50 for each experimental dollar, according to their performance. They earned $12.31 on average, including $3 for participating.

We conducted three sessions, with 10 subjects and 8 three-round games ("trials") in each, for a total of 120 observations. We will first discuss the offers players made, and how often they rejected offers. Our major conclusion is that we have closely replicated the standard findings on offers and rejections from other experiments. We then turn to the information search data which provides strong evidence of a relationship between information use and subsequent offers and acceptance decisions made by players.

4.1 Offers, rejections, and counteroffers

Three stylized facts have emerged from alternating-offer bargaining experiments (e.g., Ochs and Roth, 1989).[12] First, players do not typically choose perfect equilibrium divisions. The average offer lies somewhere between equal split and equilibrium. Second, while no offers should be rejected, some (10-20%) are, and about half of the equilibrium offers are rejected. Finally, <u>most</u> counteroffers (about 80%) are "disadvantageous." They yield <u>less</u> to the person making the counteroffer than the amount that same person previously rejected.

Figure 3 shows a histogram of offers made in the first round (pooling across sessions and trials), on the left, and a plot of rejected first round offers vs. the resulting second round counteroffer, on the right. The average offer is $2.11, about a third of the way from the equal split point of $2.50 to the equilibrium of $1.25. Most offers (88%) are closer to the equal split than to the equilibrium. However, only 13% of the offers are very close to the equal split, which suggests subjects are searching for a middle ground between the equal split and an offer which gives them more but is likely to be accepted.

Insert Figure 3 here

The shaded portion of each bar in the Figure 3 histogram indicates the number of rejected offers. Offers are rejected 10.8% of the time, a rejection rate comparable to those in earlier research. Equilibrium offers (between $1.20 and $1.40) are rare, and offers below $1.80 are rejected about half the time.

First-round rejections result in a second round of play. Of the offers made in the second round, Figure 3 shows that 85 percent of the offers are disadvantageous, falling above the diagonal line. These leave the responder (player 2) with less than he or she would have received by accepting the first-round offer from player 1. This large percentage of disadvantageous counteroffers is also comparable to prior results.

The solid circles in the right panel of Figure 3 represent second round offers that were rejected. Of all second round offers, 21 percent (3/14) are rejected, resulting in a third round. Two of the three third-round offers were rejected and resulted in no payment for either side. Given the novel information display, it is comforting to note that the results replicate the stylized facts about offers, rejections, and counteroffers observed in other experiments. To examine this more carefully, we ran one session using the same methods and parameters, but keeping the boxes open (so subjects did not have to move a mouse to acquire information). The results are a close replication: The average first round offer was $2.10, and

7.5% of these offers were rejected[13].

4.2 Information search

Player 1: Prior to opening offer.

We start our analysis of information search patterns by studying information search by player 1 subjects in the first round of each game (pooled across all eight trials of all three experimental sessions). Table II presents these average data, and Figure 4 presents the equivalent icon graph and histograms of the distribution.

Insert Table II here

Table II shows that most of the looking time (12.91 seconds out of 20.82 total seconds) is spent looking at the first-round pie size; this fact is also reflected graphically in Figure 4. Half as much time is spent looking at the second round. Subjects glance at the third box for only about one second. Notice that each box is opened about 2-4 times (i.e. the "number of acquisitions") each trial. That means subjects are opening and reopening boxes frequently, rather than memorizing the numbers in the boxes.

The pattern of transitions between boxes is shown in the last three columns of Table II and the arrows in Figure 4. In the table, entries show the average number of transitions from the row box to the column box. (For example, players moved from the round 3 box to the round 2 box about .88 times per trial.) Contrary to the backward induction prediction, there are always more forward transitions (above the diagonal) than backward ones (below the diagonal).

The last two columns in Figure 4 show histograms of the number of acquisitions and total looking times for each box. The difference in both number of acquisitions and looking time among the boxes is quite apparent, as can be seen by comparing the distributions for different rounds within each column. It is particularly striking that some subjects do not look at the second and third round payoffs at all when making first round offers. The first bars of the acquisition histograms in the middle column show the number of subjects who never open the payoff box (i.e., they open it zero times). While all respondents look at the first round payoff at least once, 19 percent never look at the round 2 payoff and 10 percent never look at the round 3 payoff, data. These subjects cannot possibly be calculating the perfect equilibrium.[14]

The relation between information search and first-round offers

It is useful to examine whether differences in information processing correlate with differences in offers. We divided trials into three group based on first-round offers. The result is shown in Figure 5. Each column represents average information search for trials in which player 1's make roughly the same offer. The left-hand "near-perfect" group are offers below $2.01 (n=38). The middle group are offers between $2.01 and $2.39 (n=56). The right-hand group represent equal split offers, between $2.40 and $2.72 (n=26). In this figure, the filled box corresponds to 14.21 seconds of looking time, while the widest box corresponds to 5.87 acquisitions.

Insert Figure 5 here

Information search in the near-perfect offer group (<$2.01) shows more backward induction reasoning than the other groups. Subjects making near-perfect offers look at the second round pie size more often (they open the box more often than the open the first-round box), and they move back and forth between the second and third rounds more than once each trial.[15] Thus, there is a strong relationship between the tendency to search for information in ways consistent with backward induction, and the tendency to make offers closer to equilibrium.

A natural hypothesis is that players learn to make perfect offers from experience. Figure 6 shows the process measures for each of the four trials in which the player made a first round offer. The figure shows almost no systematic changes with experience, and there is not significant effect on offers.[16]

Information Search by Player 2.

Insert Figure 6 here

So far we have concentrated on first-round offers made by player 1. What does information search by player 2's look like as they decide whether to accept or reject an offer? (Keep in mind that they could only open boxes <u>after</u> receiving an offer.)

Insert Figure 7 here

Figure 7 is an icon graph of the information search by Player 2s, broken down by whether offers were rejected or accepted. Acceptances are grouped by the size of the offer accepted. Overall, the search patterns look much like the patterns for player 1, with attention concentrated on the first round payoff. The players who accept small offers (<$2.00) look at the second round much longer, and make more backward transitions from the second round to the first, than players who accept larger offers.

Recall that almost all (12/14) of the rejected offers were below $2.00. Therefore, a telling comparison is between the "rejected" (first) column in Figure 7 and the accept "<2.00" (second) column in Figure 7. Those who accepted low offers looked at the second-round pie value almost twice as long as those who rejected the proposal. They also made more transitions back and forth between the first and second round pie boxes.[17] While looking ahead more is correlated with accepting a low offer, we do not know the direction of causality.

Insert Figure 8 here

Information search in the second round

How do players examine information in rounds 2 and 3? We have little data on those rounds because few first-round offers were rejected. For the few data we have (n=14), Figure 8 shows search patterns for both players in round 2. As expected, their attention shifts dramatically to the current (second) round box. One might have guessed from the first-round results that subjects always look one round ahead (a conclusion justified by the unusual data reported by Neelin, Sonnenschein and Spiegel, 1988). But Figure 8 shows that this conclusion is too simple because they do not look at the <u>third</u> round box much when they make second-round offers. Figure 8, along with Figures 4 and 7 suggest a current-round bias and a size bias: Whatever round they are in, players look at the pie size from that round most often; and players look less at small pie sizes independent of round.

Study 2: Summary

A clear picture emerges from Study 2: The offers, rejections and counter offers, and information search data show that most players do not employ a backwards induction strategy for thinking about the game (and they differ from the instructed subjects in study 1). However, some players do seem to search for information in a way consistent with backward induction, and those players are more likely to make, and accept, offers that are closer to the subgame perfect equilibrium.

While this result supports our claim that information acquisition is tied to offers, it does not entirely eliminate different explanations for the discrepancies between the game theory analysis and the observed data. Further treatments in Study 3 attempt to do so.

5. Study 3: Bargaining with Robots and Instructions

In Study 2, subjects who know how to reason game-theoretically may not do so because they believe other subjects are concerned with fairness. (Indeed, offering $1.25 is not profit-maximizing given the rejection rates observed.) Thus, while we can conclude that most subjects do not behave in a manner consistent with the subgame perfect equilibrium, we cannot determine if that is because they want to be fair (or think others do), or do not know how to calculate the equilibrium.

In Study 3, we control for fairness considerations by matching players with robot players implemented on computers, programmed to play the subgame perfect strategy. Game-theoretically inclined subjects should offer less than they did in Study 2. In addition, the robot players provide consistent feedback to the human players. Because the robots always offer the subgame perfect split, it give the participants the best chance to learn the backward induction strategy from experience. Since our procedure is otherwise is identical to that of Study 2, we can assess the relative impact of fairness vs. limited computation on departures from subgame perfect play by comparing the results.

To remove concerns about fairness, subjects were explicitly informed (see Appendix C) that they play against a computer which cares only about its own payoff, and expects the subjects to behave the same way. They were also told that trials were independent of each other, so there are no reputation or memory effects. Then the subjects played the first set of eight trials, with same equilibrium as in Study 2. This study used 30 subjects, Wharton undergraduates recruited from business classes, or from general sign-up sheets posted around the campus and procedures identical to those of Studies 1 and 2.

After finishing the first eight trials, subjects were given explicit instructions (identical to those used in the Study 1, see Appendix B) on using backward induction in these games. Two numerical examples were given, and the subjects were walked through a third example where they had to figure out the optimal offer using the backward induction strategy, and received feedback from the system. Then subjects were instructed to use this strategy when playing the remaining eight trials. They were again reminded that they were playing against the computer and the trials were independent of each other.

In the second set of eight trials, the first four trials had the same equilibrium as in the first eight trials of this study (and all previous studies). The second four trials had a different equilibrium, to assess whether subjects can generalize the strategy they learned to a different set of parameters. At the end of the session subjects were paid in cash, according to their performance. They earned $16.24 on average, including $3 for attending.

5.1 Results

Offers, Counteroffers, and Rejections.

Players in Study 3 played 8 rounds of the game with the programmed players prior to receiving instruction in backward induction. In half of the rounds they made opening offers, and in half of the rounds they received offers. If concerns for fairness are all that is preventing subgame perfect play among people, offers to robots should be near the subgame perfect equilibrium of $1.25. If generous offers to people come from limited computation, we should see generous offers to robots too.

Table III shows the offers and rejections from Study 2, aggregated over the 30 human players. Each column represents the mean offer and the fraction of rejections (out of a possible 30) for each pair of rounds, in which each player either made one first round offer or decided to accept or reject the offer made by the robot.

The offers in trials 1-8 are significantly lower than those made in Study 1, averaging $1.84 in contrast to the $2.11 in Study 1. At the same time, the observed offers are significantly above the game theoretic prediction of $1.26, and there is no clear convergence toward that level with the limited experience that is provided by the first eight rounds of this study.

Rejections are much more frequent, at least initially: More than half (61/120 or 51%) of first round offers made by the game theoretic computerized players are rejected. Rejections occur much more frequently in trials 1-2, in which 83% (25 out of 30) of all $1.25 robot offers are rejected, but only 40% of the offers are rejected in trials 3-8.

After being exposed to instruction in backward induction, there is a dramatic change. The average offers are quite near the equilibrium, $1.22, and the frequency of rejections falls to 7%. When presented with a set of pies with different parameters, in trials 13-16, players still make offers quite close, on average, to this new equilibrium and rarely reject the game theoretic offers. This finding is particularly encouraging because it means players have learned a general principle (the backward induction strategy) which transfers to a similar game with a different equilibrium.

To summarize, playing against game theoretic players does not quickly produce equilibrium play among uninstructed players. The removal of fairness considerations **did** lower offers, from $2.11 to $1.84, but did not produce play very close to the equilibrium. Fairness appears to account for only $.27 of the $.86 gap between observed offers and equilibrium.

This gap disappears after instruction in backward induction however, and generalizes to a new problem with different parameters. In sum, the offers and rejections from Study 2, show that the removal of fairness has a small but significant effect. The addition of instruction in backward induction has a much larger effect, and seems necessary to produce generalizable equilibrium play. More remarkable, perhaps, is the fact that participants neither (1) lowered their offers to robots over time, nor (2) accepted game theoretic offers. They persistently left money on the table because the robots would have accepted any offer as low as $1.25. Rejecting the robots' first-round offers of $1.25 also cost them money because second round counteroffers the robots would accept would never give them more than $1.25.

Insert Table III here

Information Search

Figure 9 shows the icon graph for all eight trials in which subjects were player 1 and made offers to robots. This data corresponds to the offers reported in the second row in Table III. As can be seen in the figure, the looking time, acquisitions, and pattern of search for the first four offers look much like that which we see for non-backward inductors, with the possible exception of the third session, which shows some tendency toward increased looking time for the third round. A major change occurs after instruction (between the 4th and 5th columns) where the first round payoff receives much less attention and more attention is to focused on the second and third round payoffs. This pattern mirrors the shift in offers that we see in the data in Table III.[18] Across the 16 trials there is a reasonable correlation, -.49, between the amount of time spent looking at the third round payoff, and the deviation from the equilibrium price. In other words, those who looked at the third round payoff tended to make offers consistent with backward induction, further evidence linking information acquisition and bargaining behavior.

Insert Figure 9 here

6. Implications and Conclusions

We studied offers and patterns of information search in three-round alternating-offer bargaining over a “shrinking pie”. As in most earlier experiments, offers were scattered between an equal split of the first-round pie ($2.50) and the perfect equilibrium offer (derived from backward induction) of $1.25. The mean offer was $2.11. About 10% of the offers were rejected.

The discrepancy between their offers and the $1.25 prediction could be due to subjects being able to use backward induction but being unsure about others’ rationality or being constrained by fairness (either they prefer to make fair offers, or worry that others will reject unfair offers), or by subjects not using backward induction. We tested these different hypotheses in two ways.

First, we conducted sessions in which players bargained with self-interested robots programmed to play the perfect equilibrium strategy. Subjects who do backward induction instinctively, but are constrained by fairness or doubts about rationality when they bargain with other people, should make $1.25 offers to the robots. They offered only $1.84 on average, which establishes that most subjects do not do backward induction instinctively.

Second, we measured patterns of information search by showing the three rounds of `pies’ in boxes which could only be opened by moving a cursor into a box. These data show that most subjects did not look at the pie sizes in the correct order, and for the length of time, necessary for backward induction (compared to a control group trained in backward induction). Instead, most subjects concentrated on the current round when making decisions and looked ahead insufficiently. In 19% and 10% of the trials, respectively, they did not even open the second and third-round boxes. Subjects who looked ahead more often did make offers closer to the equilibrium prediction, and were more likely to accept such offers.

The comparison of human and robot bargaining, and data on information search, together show that fairness and limited use of backward induction both play a role. Fairness matters because about a third of the gap between the prediction and the mean offer was eliminated when subjects switched from bargaining with other subjects to bargaining with robots. (Many experiments on simple dictator and ultimatum bargaining games also show that fairness matters.) However, the fact that subjects do not search for information by looking ahead, and the fact that a group trained in backward induction do make perfect equilibrium offers, shows that most subjects do not use backward induction instinctively. Furthermore, since offers and information search do not change across several trials, subjects do not learn much from experience.

The lack of foresight our subjects demonstrated is evident in many other settings, including the “dollar auction” (Shubik, 1971), asset markets exhibiting speculative bubbles (e.g., Porter and Smith, 1995), and repeated prisoners' dilemmas or (closely-related) centipede games (McKelvey and Palfrey, 1992). These results could be generated by subjects who can backward induct, but aren’t sure that others can. Our results allow us to separate the two explanations, and support the notion that failures of backward induction are part of the story. In further research, the use of robot players might help disentangle fairness, lack of mutual knowledge of rationality, and limited computational accounts of these phenomena and others.

The results raise an interesting question: If subjects do not use backward induction, how are they computing offers? We think they make heuristic judgments about which future nodes are likely to be reached-- ignoring more distant nodes, or lower-payoff nodes--, eliminate those from the start, and reason vaguely game-theoretically about the simplified game. Note that subgame perfection can also be interpreted algorithmically as a special, labor-intensive heuristic which eliminates future nodes, by figuring out which nodes are likely to be reached by working through all of them.

One interpretation of our results is that we measure strategic heuristics which were evolutionarily adapted to facilitate social exchange in primitive environments and are encoded in specialized cognitive modules. Presumably, this adaptation did not equip people to do backward induction, but did equip them to react hostilely to unfair offers, and to anticipate hostile reactions of others. We see our research and this evolutionary view as complementary. We are simply trying to understand how subjects think when they bargain, by measuring what information they look at it, in order to understand which offers they make and accept. The evolutionary interpretation jumps further back in time, asking where the bargaining heuristics we observe might have come from.

It would be useful to measure information processing in other games in which game-theoretic principles make implicit predictions about how information use affects choices. For example, Costa-Gomes, Crawford, and Broseta (1997) test whether players look at game payoffs in ways consistent with iterated eliminated of dominated strategies. Another example is forward induction. Consider a two-stage game in which player 1 has a preplay option in the first stage. If she rejects the option, players 1 and 2 play a simultaneous-move game in the second stage. Forward induction predicts that if the second stage is reached, player 2 will look back at player 1's foregone payoff to figure out what player 1 intends to do in the second stage. Experimental evidence that people do forward induction is mixed (e.g., Cachon and Camerer, 1996). In a study measuring information search, Cachon, Camerer and Johnson (1994) show that violations of forward induction are common and can be predicted by how frequently player 2 looks back at player 1's foregone payoff.

Failures of forward and backward induction are two examples of how cognitive limits cause players to simplify their mental representations of games and use heuristics rather than complicated algorithms to choose strategies. More generally, attention is a scarce resource and allocating it optimally is necessary to compute a game-theoretic equilibrium. By measuring attention directly, this paper contributes to further understanding of both the successes and failures of game theoretic predictions and the roles of thinking and learning in influencing these outcomes. Furthermore, our emphasis is consistent with the sensible shift away from interpreting game-theoretic equilibria as solutions that brilliant players figure out, toward thinking of equilibria as resting points which are the result of evolution or learning by players of limited rationality (e.g., Milgrom and Roberts, 1990; Weibull, 1995; Camerer and Ho, in press). This shift parallels studies of individual choice with examine how heuristic processes can be procedurally rational (Payne, Bettman and Johnson, 1993).

REFERENCES

Anderson, J.R., Pirolli, P., and Farrell, R. (1988). "Learning to program recursive functions," in *The Nature of Expertise* (M. Chi, R. Glaser, and M. J. Farr, Eds.). Hillsdale, NJ: Erlbaum.

Binmore, K., Rubinstein, A., and Wolinsky, A. (1986). "The Nash bargaining solution in economic modelling," *RAND Journal of Economics*, **17**, 176-188.

Binmore, K., Shaked, A., and Sutton, J. (1985). "Testing noncooperative bargaining theory: A preliminary study," *American Economic Review* **75**, 1178-1180.

Binmore, K., Shaked, A., and Sutton, J. (1988). "A further test of noncooperative bargaining theory: Reply," *American Economic Review* **78**, 837-839.

Bolton, G.E. and A. Oeckenfels (1997). "ERC: A theory of equity, reciprocity, and competition," Penn State University working paper.

Cachon, G., and Camerer, C.F. (1996). "Loss-avoidance and forward induction in experimental coordination games," *Quarterly Journal of Economics* **444**, 165-194.

Cachon, G., Camerer, C.F., and Johnson, E.J. (1994). "Information processing and forward induction," University of Pennsylvania Department of Decision Sciences Working Paper.

Camerer, C.F. (1998). *Experimental Studies of Game Theory*. Book manuscript in progress.

Camerer, C.F., and Ho, T.H. (In press). "Experience-weighted attraction learning in normal-form games," *Econometrica*.

Camerer, C.F., Johnson, E.J., Sen, S., and Rymon. T. (1993). "Cognition and Framing in Sequential Bargaining for Gains and Losses" in *Frontiers of Game Theory* (K. Binmore, A. Kirman, and P. Tani, Eds.), MIT Press.

Camerer, C.F., and Thaler, R.H. (1995). "Anomalies: Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives*, **9**, 209-219.

Card, S.K., Moran, T.P., and Newell, A. (1983) *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Erlbaum.

Cox, J.C., and Oaxaca, R.L. (1989). "Laboratory Experiments with a Finite-Horizon Job- Search Model," *Journal of Risk and Uncertainty*, **2**, 301-329.

Costa-Gomes, M., Crawford, V, and Broseta, B. (1998). "Cognition and behavior in normal-form games: An experimental study," University of California, San Diego Department of Economics Working Paper.

Eckel, C., and Holt, C.A.(1989). "Strategic voting in agenda-controlled committee experiments," *American Economic Review* **79**, 763-773.

Fehr, E., and Schmidt, K. (1997). A theory of fairness, competition, and cooperation. University of Zurich Working Paper.

Forsythe, R., Palfrey, T., and Plott, C. (1982). "Asset valuation in an experimental market," *Econometrica* **50**, 537-567.

Güth, W., and Tietz, R. (1990). "Ultimatum bargaining behavior: A survey and comparison of experimental results," *Journal of Economic Psychology* **11**, 417-449.

Harrison, G.W., (1989). "Theory and misbehavior of first-price auctions," *American Economic Review* **77**, 749-762.

Harrison, G.W., and McCabe, K. (1992). "Testing bargaining theory in experiments," in *Research in Experimental Economics* (R. M. Isaac, Ed.), Vol.5. 1992.

Harrison, G., and McCabe, K. (1996). "Expectations and fairness in a simple bargaining experiment," *International Journal of Game Theory* **25**, 303-327.

Johnson, E.J., Payne, J.W., and Bettman, J.R. (1988). "Information displays and preference reversals". *Organizational Behavior and Human Decision Processes* **42**, 1-21.

Johnson, E.J., and Schkade, D.A. (1989). "Heuristics and bias in utility assessment: Further evidence and explanations," *Management Science* **35**, 406-424.

Kahneman, D., and Tversky, A. (1979). "Prospect Theory: An Analysis of decision under risk," *Econometrica* **47**, 263-91.

Kennan, J., and Wilson, R. (1990). "Can strategic bargaining models explain collective bargaining data?," *American Economic Review* **80**, 405-409.

Levine, D. K. "Altruism and spite in bargaining experiments", UCLA Department of Economics working paper, 1997.

Loewenstein, G., Thompson, L., and Bazerman, M. (1989). "Social Utility and Decision Making in Interpersonal Contexts," *Journal of Personality and Social Psychology* **53**, 426-441.

McKelvey, R.D., and Palfrey, T.R. (1992). "An Experimental Study of the Centipede Game," *Econometrica* **60**, 803-836.

Neelin, J., Sonnenschein, H., and Spiegel, M. (1988). "A further test of noncooperative bargaining theory: Comment," *American Economic Review* **78**, 824-836.

Newell, A., and Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R.E., and Wilson, T.D. (1977). "Telling more than we can know: Verbal reports on mental processes," *Psychological Review* **84**, 231-259.

Ochs J. and Roth, A.E. (1989). "An experimental study of sequential bargaining," *American Economic Review* **79**, 355-384.

Osborne, M.J., and Rubinstein, A. (1990). *Bargaining and Markets*. San Diego, CA: Academic Press.

Payne, J.W., Bettman, J.R., and Johnson, E.J. (1988). "Adaptive strategy selection in decision making," *Journal of Experimental Psychology: Learning, Memory, and Cognition* **14**, 534-552.

Porter, D.P., and Smith, V.L. (1995). "Futures contracting and dividend uncertainty in experimental asset markets," *The Journal of Business* **68**, 509-541

Rabin, M. (1993). "Incorporating Fairness into Game Theory and Economics," *American Economic Review* **83**, 1281-1302.

Roth, A.E. (1995). "Bargaining experiments," in *Handbook of Experimental Economics* (G.E. Roth and J. Kagel Eds.). Princeton: Princeton University Press.

Rubinstein, A. (1982). "Perfect equilibrium in a bargaining model," *Econometrica* **50**, 97- 109.

Russo, J.E. (1979). "A software system for the collection of retrospective protocols prompted by eye fixations," *Behavior Research Methods & Instrumentation* **11**, 177-179

Schkade, D.A., and Johnson, E.J. (1989). "Cognitive processes in preference reversals," *Organizational Behavior and Human Decision Processes* **44**, 203-231.

Shubik, M. (1971). "The dollar auction game," *Journal of Conflict Resolution* **15**, 109-111.

Spiegel, M., Currie, J., Sonnenschein, H., and Sen, A. (1990). "Fairness and strategic behavior in two- person, alternating-offer games: Results from bargaining experiments," *Games and Economic Behavior*.

Ståhl, I. (1973). "Bargaining Theory," *Swedish Journal of Economics* **75**, 410-413.

Weibull, J. (1995). *Evolutionary Game Theory*. Cambridge: MIT Press.

Appendix A: Instructions

This is an experiment about bargaining. You will bargain using the computer in front of you, in several sessions. In each session you will either be a BUYER or a SELLER. Each BUYER will be randomly paired with a SELLER in a session. You will be paired with a different person in each session.

Please go to the next screen

At the end of the experiment, you will be paid half of the total amount you actually earned bargaining, plus $3 for attending.

Please do NOT talk to other subjects at any point during the experiment. Direct any questions you may have to us.

Go to the next screen....

This is how the bargaining works: In each session the bargainers will divide an amount of money, which we will call a "pie". The pie is initially worth some amount ($4.00, for example). Each session has three rounds. In the FIRST round, the SELLER will offer part of the pie to the BUYER; the rest of the pie is left for the SELLER. If the SELLER's offer is accepted by the BUYER, the session ends and the bargainers earn their shares of the pie.

Go to next screen....

If the SELLER's offer is rejected by the BUYER, the session continues to a SECOND round. The SECOND round is like the FIRST except the pie is smaller (for example, $2.00 instead of $4.00) and the roles of BUYER and SELLER are reversed: the original BUYER makes an offer to the original SELLER. If the BUYER's offer is accepted by the SELLER, the session ends. If the offer is rejected, the session continues to a THIRD round.

Go to the next screen....

In the THIRD round, the pie will get even smaller (for example, $1.00 instead of $2.00) and the roles are reversed again: the original SELLER makes an offer to the original BUYER. If the SELLER's offer is rejected in the THIRD round, the session ends (there is no fourth round), and both bargainers get zero in that session.

Notice that the pie in the SECOND round will be split only if the FIRST round goes on. And similarly, the pie in the THIRD round will be split only if the SECOND round goes on.

Go to the next screen....

Now we will illustrate how the bargaining works through an example. It is shown on a sheet in front of you. Please wait here until we go through the example.

Thank you in advance for your participation.

Now, we will walk you, screen by screen, through an EXAMPLE session using the computer and the mouse.

All the sessions in the actual experiment will have the same structure as this EXAMPLE

Go to next screen....

The EXAMPLE is now over. The experiment starts with the next session.

Please don't feel that any particular kind of behavior is expected of you. There is no correct way of bargaining in this experiment. All your responses will be coded by a subject number, so your responses will be anonymous to us.

So that you can keep track of the bargaining process in each session, please record offers and accept/reject responses and calculate your share of the pie in all three rounds of each session on the worksheet provided.

However, please do not write down total pie values on your worksheet.

Remember: Do NOT talk to other subjects at any point during the experiment. Direct any questions you may have to us.

Appendix B

The first part of the experiment is now over. Please wait here until the experimenter tells you to go to the second part.

One way of bargaining in the situation you were just in is to think first about what would happen in the third (last) round of the session. Imagine that you are a seller in the third round. You would make an offer to the buyer. If she or he agreed to the split, they would get what you offered them and you would get the remainder of the pie.

If the buyer were trying to make the most money and knew that you were trying to do the same, then the buyer should settle for any amount greater than zero (which is what they would receive for that session if they rejected your third round offer). The buyer will know that the most he can get in the third round is a penny, so you can get the third round amount minus a penny.

Now, in the second round, the buyer should know what you could make in the third round. In the second round, he could offer you one penny more than what you could make in the third round, and you would accept it. The buyer would get the rest of the second round pie (the total pie minus the amount you would accept).

Now imagine that you are in the first round of the session. You know what the buyer can get in round 2. In round 1, you should offer him or her one penny more than what she or he could make in round 2.

An example may help in understanding the bargaining strategy that was just described to you. Imagine that in a particular session the first round pie value is $4, the second round pie value is $2 and the third round pie value is $1. If you were making a first round offer, what would your offer be?

If both you and the buyer were trying to make the most money, then in the third round you would offer the buyer one cent and keep 99 cents for yourself. The buyer would accept the one cent because it is greater than nothing which is what they would receive for that session if they rejected your third round offer.

Likewise, in the second round, the buyer would offer you one dollar (99 cents plus an additional cent) and keep the remaining dollar for him or herself. You would accept this offer because it would be one cent more than what you could get if you rejected this offer and went to the third round of the session.

In the first round, then, you would offer the buyer $1.01. They should accept this since it is one cent more than what they could make in round 2.

Here's another example. Imagine that the first round pie value is $3, the second round pie value is $1 and the third round pie value is 75 cents. If you were making a first round offer, what would your offer be?

In the third round you would offer the buyer one cent and keep 74 cents for yourself.

In the second round, the buyer would offer you 75 cents (74 cents plus an additional cent) and keep the remaining 25 cents.

In the first round you would offer the buyer 26 cents. You would keep $2.74 for yourself.

Before we start the second part of the experiment, we will walk you through an EXAMPLE session. Try to figure out the optimal offer using the strategy just described.

The EXAMPLE is now over. The optimal offer in this example was $2.01. If you have any questions about the example, please raise your hand and the experimenter will come to you.

The second part of the experiment starts with the next session. In the next few sessions, use the strategy we just explained to you in generating your offers and in deciding whether to accept or reject offers in each session. As before, you will be playing against a computer, and the sessions are independent of each other.

Appendix C

In generating your offers, or deciding whether to accept or reject offers, assume the following:

1. You will be playing against a computer which is programmed to make as much money as possible for itself in each session. The computer does not care how much money you make.

2. The computer program expects you to try to make as much money as you can, and the program realizes that you have been told, in instruction (1) above, that it is trying to earn as much money as possible for itself.

3. The sessions are independent of each other. The computer program does not store information about the previous sessions; it uses the same program in each session.

Endnotes

[1] In Camerer et al (1993) we studied an expanding-loss game which gives the same wealth outcomes as the shrinking-gain games studied in this paper. Offers were more dispersed and rejected more frequently, and there were differences in looking times. This difference between shrinking-gain and expanding-loss games is consistent with gain-loss distinctions or “framing effects” reported in many choice experiments (e. g., Kahneman and Tversky, 1979).

[2] In the third round, player 1 offers \$.01 or nothing (depending on whether player 2 will accept nothing) and earns \$1.24-\$1.25. In the second round, player 2 must offer \$1.24-\$1.26 to player 1, leaving \$1.24-\$1.26 to herself. In the first round, player 1 must therefore offer \$1.24-\$1.27 to player 2. For simplicity, we refer to this range of perfect equilibria as “the” equilibrium at \$1.25, recognizing that the equilibrium is not unique, and is subgame perfect (<u>every</u> offer is a Nash equilibrium). Note that a short cut to calculating equilibria in games of this form is that the first player earnings is the sum of the pie sizes in odd-numbered rounds (\$5.00 + \$1.25), minus the pies values in even-numbered rounds (-\$2.50).

[3] Binmore, Shaked and Sutton (1985) studied two-stage bargaining with pie sizes of 100 and 25. The equilibrium offer is 25, and players initially offered between 25 and 50. When player 2's were asked to take the player 1 role, in a (hypothetical) repetition of the two-round game, they offered only 25. Experience in the player 2 role seemed to teach subjects how much bargaining power player 1 has, and they were eager to exploit it in future games. Harrison and McCabe (1994) had subjects play a series of three-round games with pie sizes \$5, \$2.50, \$1.25 (as in our initial example). Between each three-round game subjects played a two-round game with pie sizes \$2.50 and \$1.25. Notice that the two-round game is simply a subgame of the three-round game (corresponding to its second and third rounds). With experience, subjects learned from playing the two-round game what player 2 would get at the second round of the three-round game (around \$1.25). Then they gradually converged near to the equilibrium division in the <u>first</u> round of the three-round game.

[4] For example, if a subject in the player 2 role rejects a near-equilibrium robot offer of \$1.50, thinking it unfair, she will end up making less since a second-stage counteroffer giving her more than \$1.25 will be automatically rejected by the robot. Harrison and McCabe (1996) also used robots in ultimatum games (unbeknownst to their subjects) to see whether repeated exposure to game-theoretic play would reduce subjects’ offers (it did).

[5]Applications include preference reversals (Johnson, Payne, and Bettman, 1988; Schkade and Johnson, 1990), reactions of decision-makers to time pressure (Payne, Bettman and Johnson, 1988), and differences in risk attitudes across different response modes (Johnson and Schkade, 1989).

Computerized measurement of information acquisition has several advantages over other methods. For instance, subjects can be tape-recorded giving on-line or retrospective explanations of their thinking (“verbal protocols”). But protocols are often too coarse to capture subtleties in thinking, they can distort the process being measured because speaking affects thinking (Russo, 1979), and subjects may be unable to explain why they make the choices they do. Another method is photographing the eyes of subjects while they look at a display of information. The pattern of eye movements reveals what information is being looked at, for how long. (Eye movements are widely used in studies of reading comprehension.) But eye movement machines are very expensive and the method is uncomfortable. Compared to these methods, MOUSELAB is easier because the mouse is a quick, low-cost pointing device which is easy to learn, unobtrusive, and more adaptable to group decision making.

MOUSELAB is also easily adapted to group situations, unlike verbal protocols and eye movements. The recording of information acquisition is automatic, so experiments can be run on networked computers without intervention by experimenters.

[6]An important assumption in our analysis is that information search is closely linked to information usage (i.e., subjects do not memorize the box contents). This assumption is plausible for four reasons. First, after a little practice using the mouse is <u>easier</u> than conscious memorization. Second, players opened some boxes several times before making a decision, suggesting that they did not memorize the boxes' contents. (We also did not observe subjects taking notes, although they could). Third, we used slightly different pie sizes in each trial to inhibit memory. The pie sizes were actually \$5+a, \$2.50+b, and \$1.25+c (where a, b, and c ranged from -\$.25 to +\$.25). By choosing a-b+c=0, the pie sizes changed each period but the equilibrium share for player 1 was held fixed. Fourth, the results show that the amount of time boxes are open does not decline much over trials, which it would if they were memorizing.

[7]There is ample evidence from cognitive psychology that the backward induction assumptions underlying the theory are likely to be false. Search through a problem space is usually limited to a few levels and a subset of possible actions (Newell and Simon, 1972). Anderson et al (1987) found that computer programmers had terrible difficulty learning the LISP language, when it involved recursion like backward induction. In experimental economics, Forsythe, Palfrey and Plott (1982), Eckel and Holt (1989), and others have found that people do not backward induct instinctively, but can learn to, in markets and games. It is notable that Cox and Oaxaca (1989, 1990) report good performance by subjects in a job-search experiment which requires backward induction (but for which heuristic search paths approximate optimal paths fairly closely).

[8] Some subjects were high school juniors attending a summer program at Penn; they behaved like college students in the lab. Spiegel, et al (1990) found few subject pool differences in bargaining game much like ours.

[9] The discrepancy between the average response \$1.22 and the equilibrium (\$1.25) is probably due to the MOUSELAB response scale, which recorded responses to the nearest nickel. All differences that we report were tested with an Analysis of Variance (ANOVA) and reached or exceeded conventional levels of significance.

[10] To distinguish between players actually examining information from those who opened boxes accidentally (usually while moving from one box to another), we filtered out all information acquisitions lasting less than .18 seconds. People do not accurately perceive anything they see that briefly (Card, Moran and Newell, 1983).

[11]Of course, to calculate an equilibrium offer to player 2, player 1 does not actually need to look at the first round payoff at all, as long as she knows the first round pie is bigger than the second round pie, as subjects were told.

[12] There are many qualifications to these stylized facts. The most important is that learning may create convergence to equilibrium, reducing the number of rejections and disadvantageous counteroffers (as in Binmore, Shaked and Sutton, 1985, and Harrison and McCabe,1994). But empirical evidence of learning is mixed.

[13] These figures are not significantly different from those in the closed-box study ($t(154)=.18$, $p>.60$, two-tailed test for offers, $z(156)=.63$ $p>.52$ for proportion of rejections). Forcing subjects to open boxes does not seem to affect the offers they make.

[14]An alternative explanation for the concentration of attention on first round payoffs is that information at the top of the display will be looked at more frequently and longer than information at the bottom, and the first round happens to be at the top. To control for this effect, we ran another study, identical to study 2, with the same instructions, computer display, pie sizes, and subject pool, except the order in which the pie sizes was displayed was inverted. The top line showed information for round 3, the middle line for round 2, and the bottom line for round 1. There were no significant differences in the proportion of time spent on each round size. However, subjects did take have longer looking times for all three boxes, presumably because of the unnatural format, and the average offer was $1.94, significantly less than the $2.11 in study 2. Thus, the most important findings-- average offers are far from equilibrium, and subjects look at the first round pie size most often, and the third round least-- are robust to flipping the box order upside down.

[15]Hypothesis tests were conducted using an ANOVA for each of the process measures, and subsequent t-tests to test for differences among the three response groups. The process measures are significantly different across the three groups, mostly because the near-perfect group is significantly different from the other two. The only exception to this pattern is the time spent on the first round payoff, which is not significantly different among the groups. In all significance tests we report, differences among groups is tested by an **F** test that the group means differ, **p** < .05. Specific differences between pairs of groups are tested by **t** tests, **p** < .05 for a priori hypotheses, and Bonferroni confidence intervals for post-hoc tests.

[16]ANOVA's on each of the process measures, similar to those above, show no differences among the trials for any of the process measures, except the number of transitions between the first and third round payoff. Those transitions are rare (occurring less than once per trial) and of no theoretical

significance.

[17]Significance tests of the differences between accepted and rejected offer processing lack power because there are so few rejections--only 14 in our 120 observations.

[18]A series of ANOVAs conducted on all the information search measures that we collected shows many significant differences when contrasting sessions 1-8 with sessions 9-16. These include significant increases ($p < .001$) in looking time for the second and third round payoff, increases in the number of acquisitions of the second and third round payoff, and increases in the number of transitions between the second and third round payoff. There were corresponding decreases in the time spent looking at the first round payoff, and the number of acquisitions of the first round payoff.

Table I: Predicted and observed offers.

Theory	**Treatments: Bargaining with...**		
	...other subjects.	...robots	....robots and instruction.
GT	$1.25	$1.25	$1.25
GT-Fairness	[1.25,2.50]	1.25	1.25
LC	p_{LC} θ [0,2.50]	p_{LC}	1.25
LC-Fairness	$p_{LCF}(s) > p_{LC}$	$p_{LCF}(r) < p_{LCF}(s)$	1.25
Actual mean offers	$2.11	$1.84	$1.22

Table II: Acquisitions, Looking Time and Transitions by Round.

Round	Number of Acquisitions	Total Time Examining Payoff	Transitions[a]		
			1	2	3
Round 1 ($5.00)	4.38	12.91	--	2.55	.65
Round 2 ($2.50)	3.80	6.67	2.10	--	1.24
Round 3 ($1.25)	2.12	1.24	.50	.88	--

[a]Numbers above the diagonal represent transitions from earlier to later rounds, those below the diagonal represent transitions from later to earlier rounds.

Table III: Offers and Rejections, Study 3.

	No Instruction, Equilibrium= $1.26				**Post Instruction, Equilibrium= $1.26**		**Post Instruction, Equilibrium= $1.51**	
Round	**1-2**	**3-4**	**5-6**	**7-8**	**9-10**	**11-12**	**13-14**	**15-16**
Average Offer	1.83	2.02	1.67	1.82	1.21	1.23	1.51	1.54
Percentage Rejections (Based on 30)	83.3	36.6	40.0	43.3	3.3	13.3	20.0	0.0

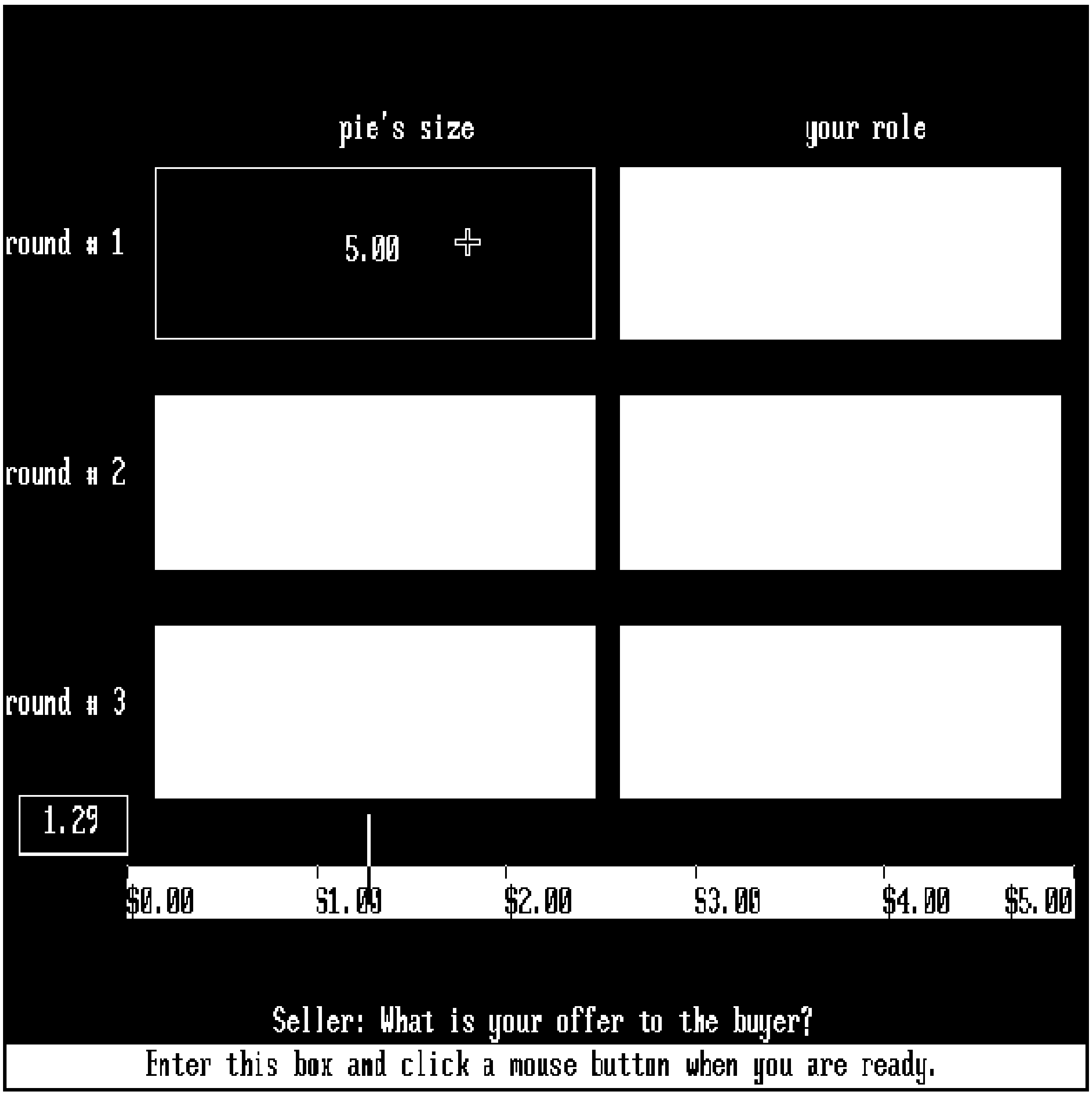

Figure 1: Computer Display Seen by Subjects

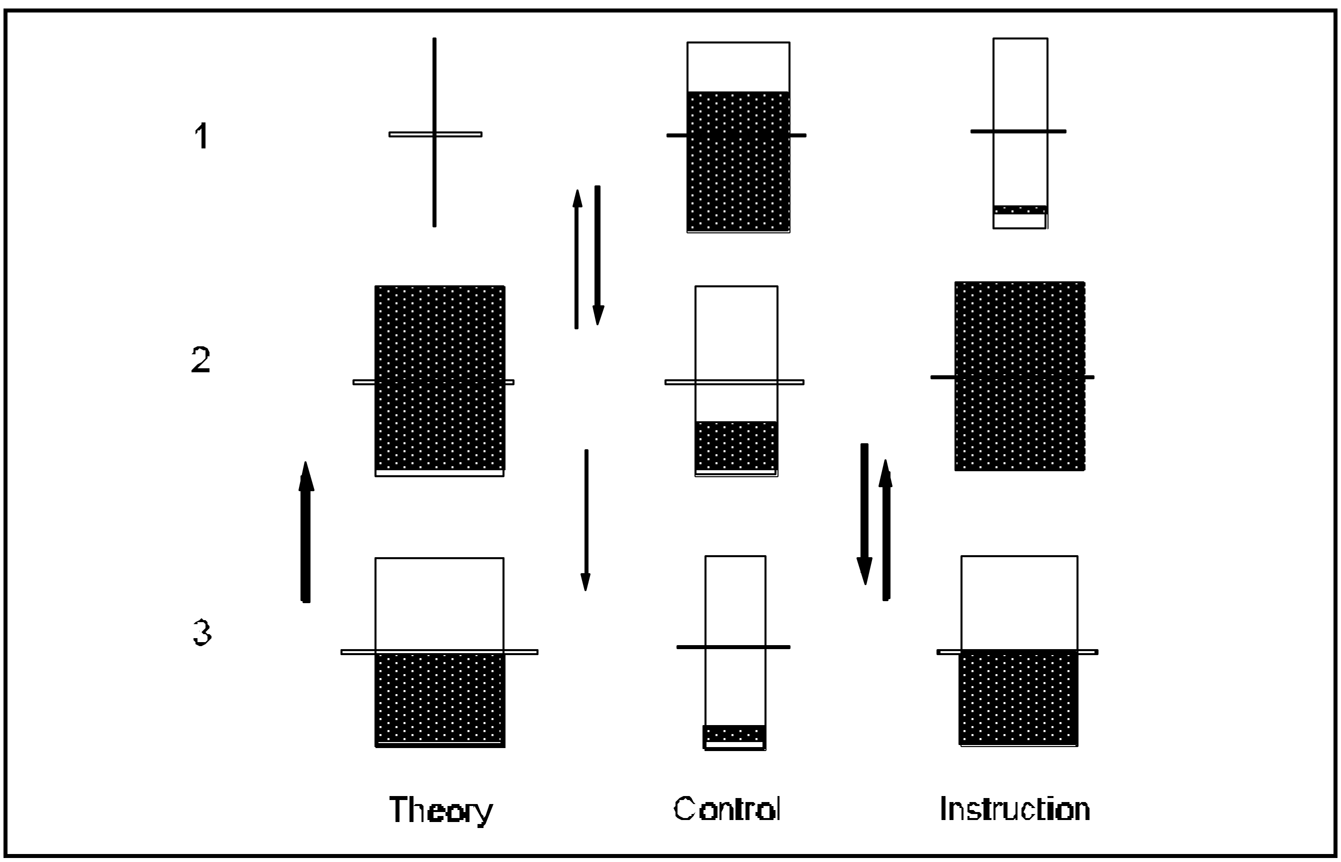

Figure 2: Theoretical Predictions and Observed Information Acquisition for Calibration Study.

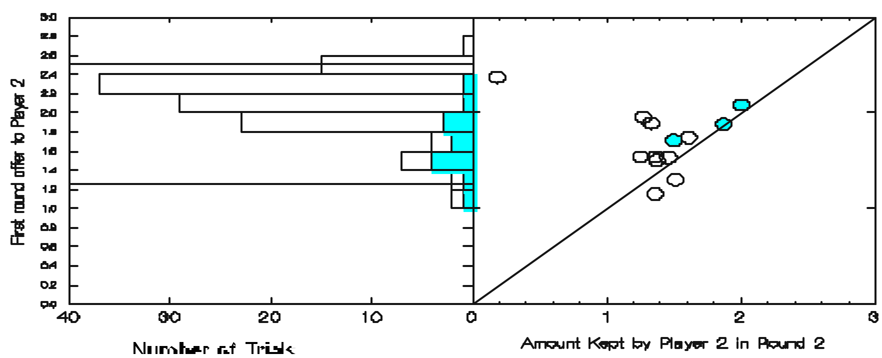

Figure 3: Distribution of Offers and Rejections.

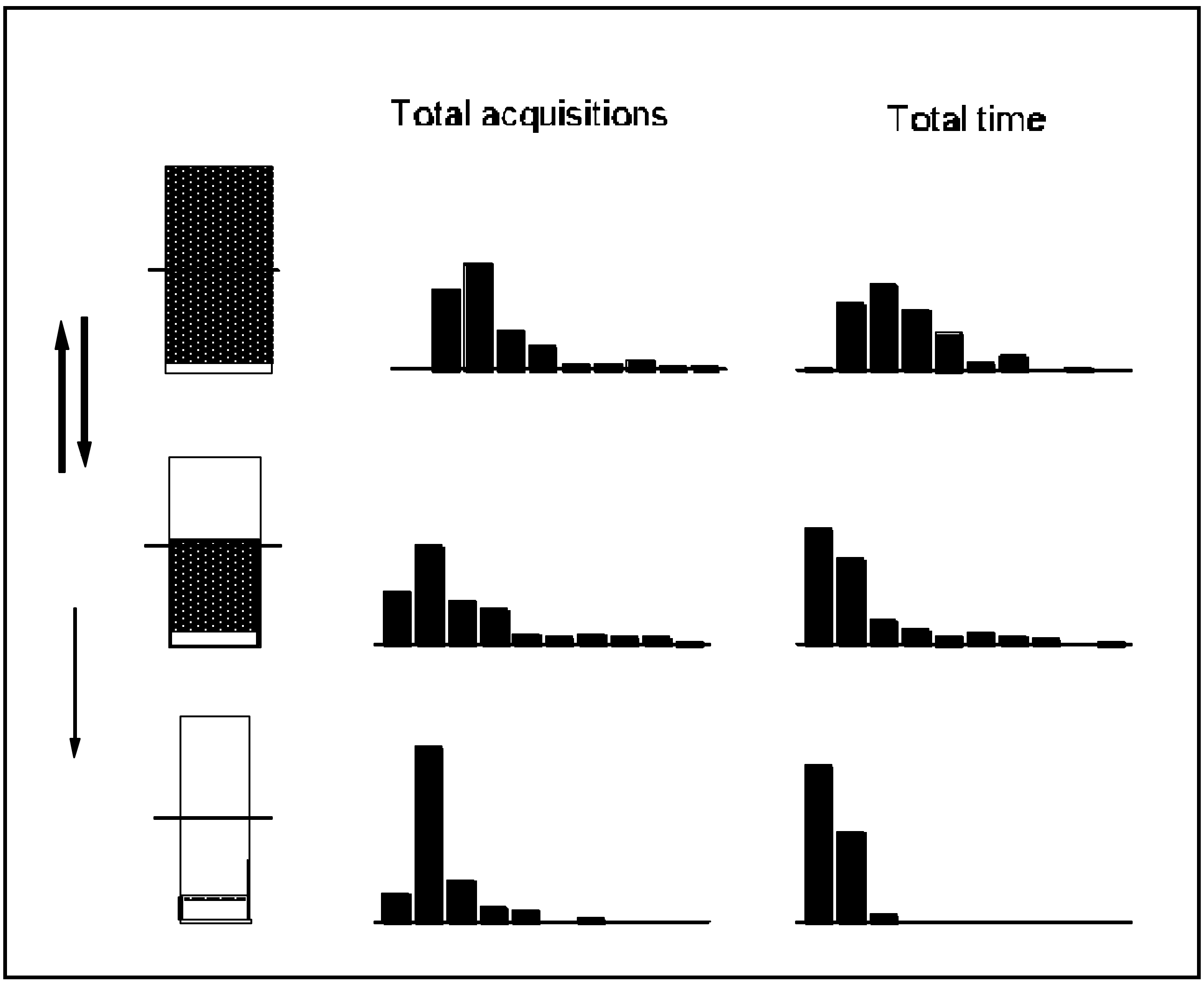

Figure 4: Number of acquisitions, total looking time and transitions for Player 1, before making a first round offer.

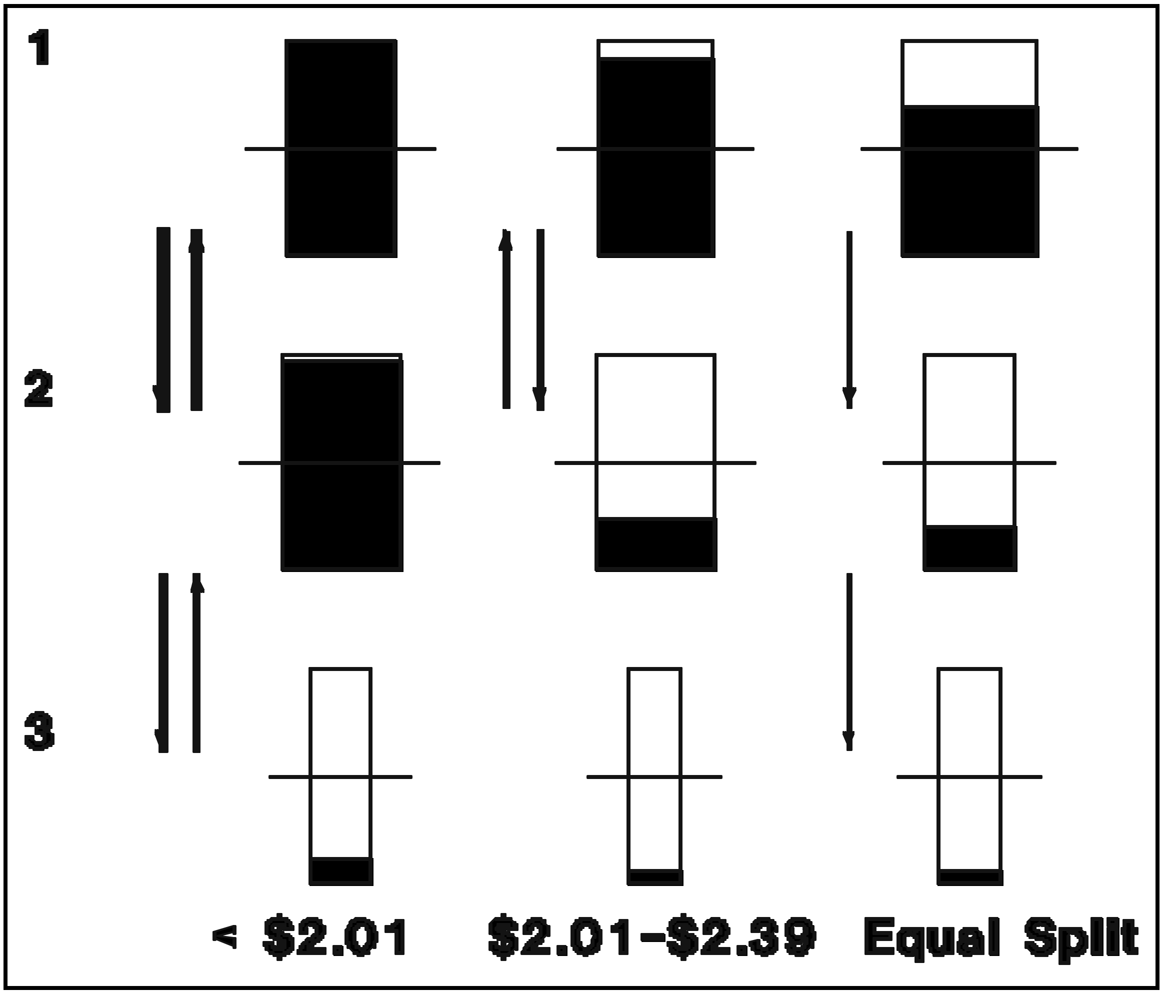

Figure 5: Number of acquisitions, looking time, and transitions by first round offer.

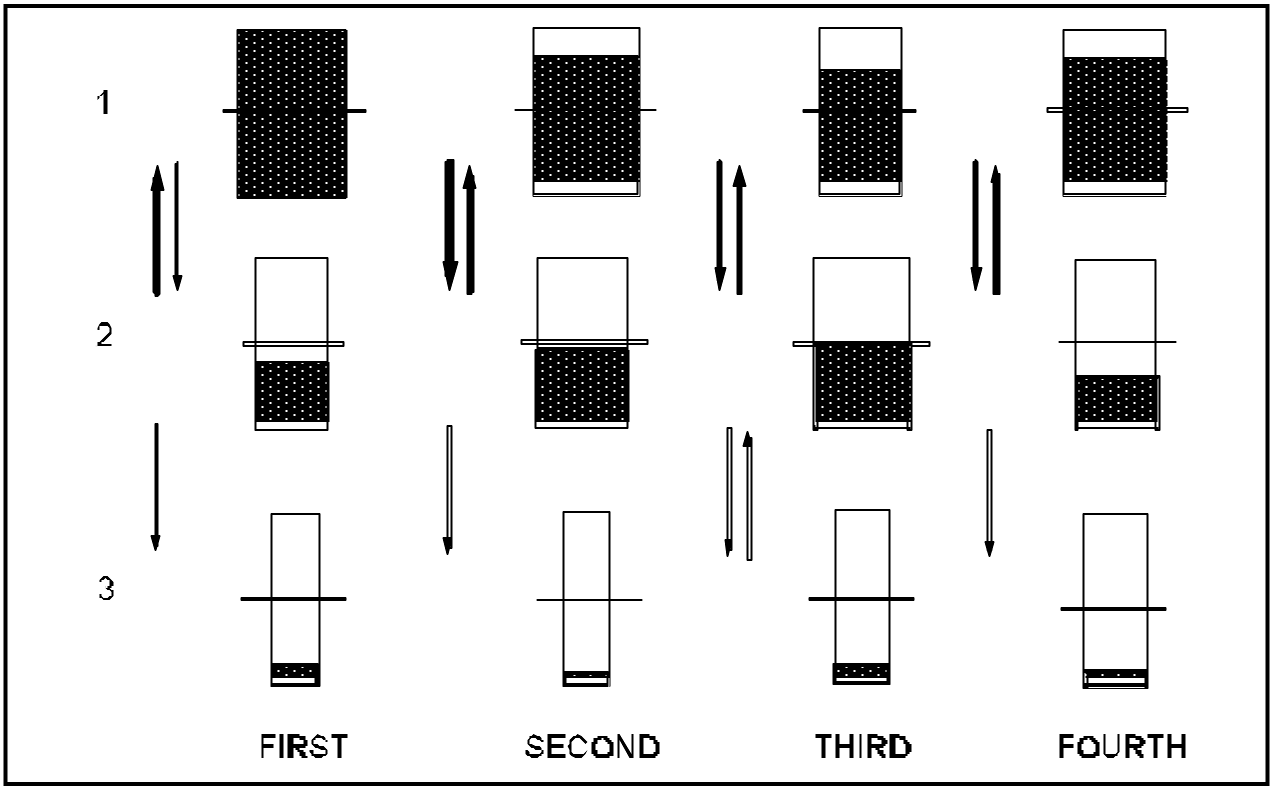

Figure 6: Number of Acquisitions, Looking Time, and Transitions by Trial

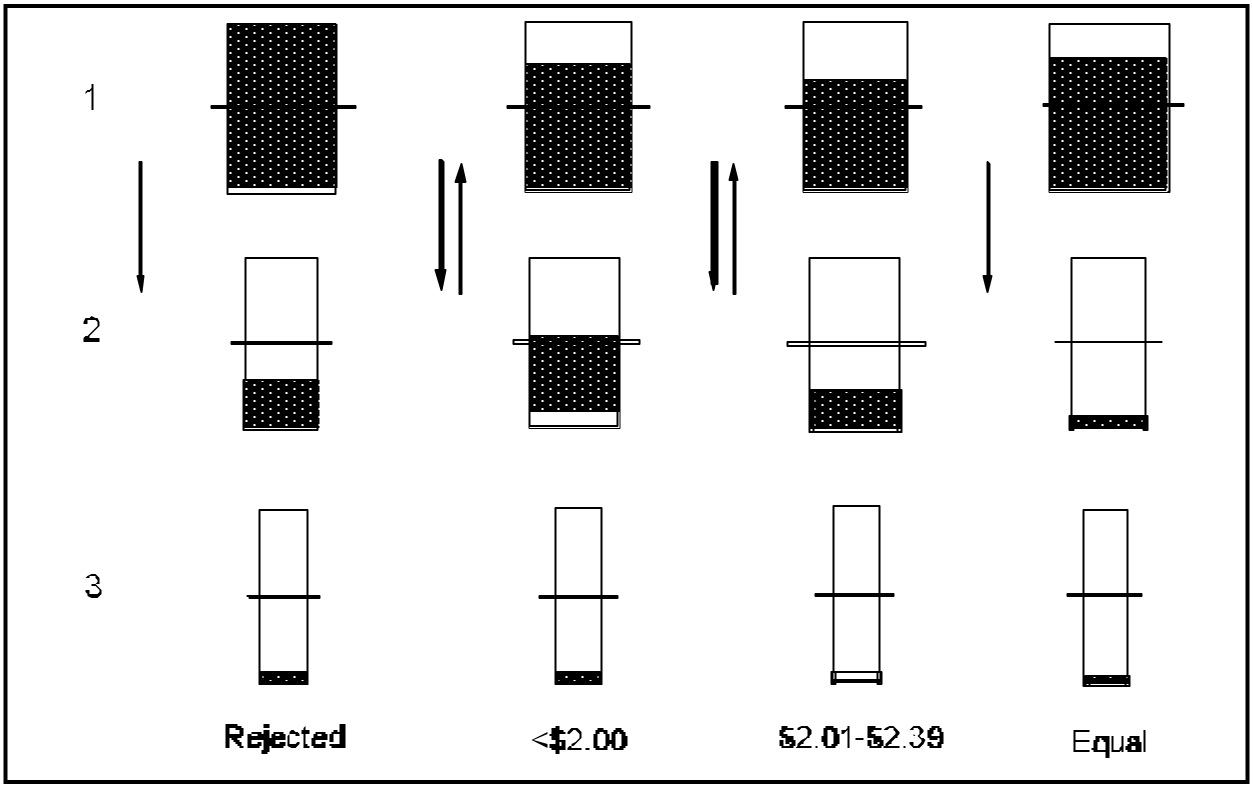

Figure 7: Information search by player 2's who accept and reject first-round offers.

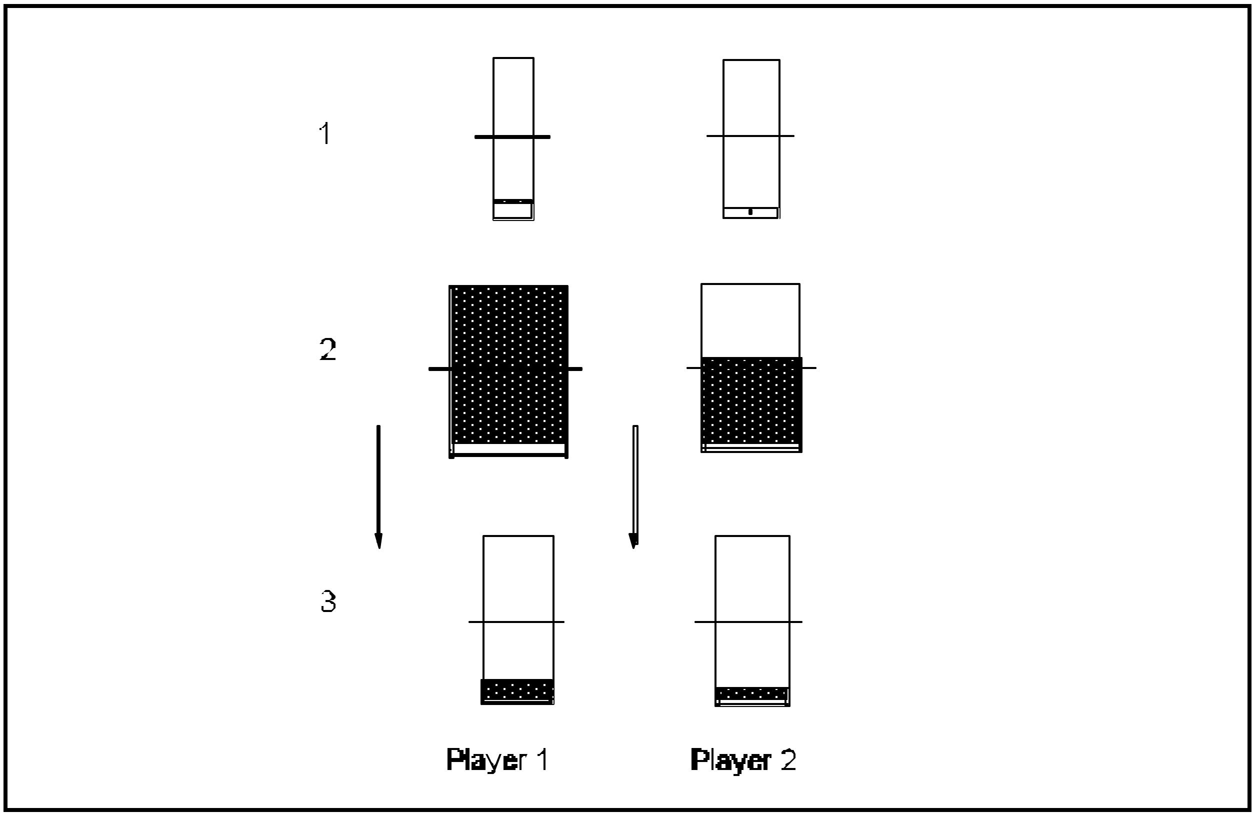

Figure 8: Information search, second round

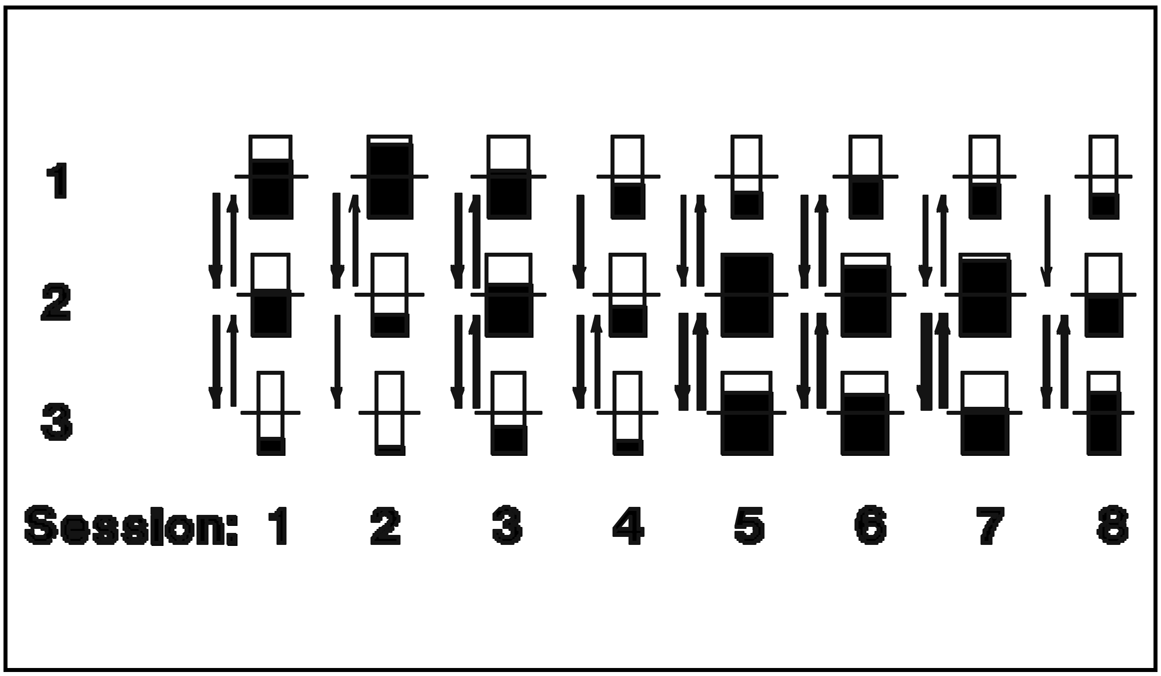

Figure 9: Information Acquisition, by session, Study 3